Int. Journal of Business Science and Applied Management, Volume 6, Issue 3, 2011

A comparison of push and pull production controls under machine breakdown

Joshua Prakash
School of Mechanical Engineering, Universiti Sains Malaysia
(USM), Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia
Telephone: +006017-3655113
Email: joshuaprakash@gmail.com

Chin Jeng Feng
School of Mechanical Engineering, Universiti Sains Malaysia
(USM), Engineering Campus, 14300 Nibong Tebal, Penang, Malaysia
Telephone: +00604-5996365
Email: chinjengfeng@eng.usm.my

Abstract

Production control for high-mix production remains a complex issue. Common pull system replenishment generates large works-in-process (WIPs) for each part type, especially under breakdown. This paper attempts to solve this by presenting a production control that classifies parts into two categories. The performances of three production control mechanisms under breakdown are compared. The production control mechanisms in consideration are push, shared constant WIP (CONWIP), and parallel CONWIP. A full-factorial simulation experiment was conducted. ANOVA was performed to determine the significant effects of input factors. Response surface methodology was used to demonstrate the behavior of performance measures in terms of these significant input factors. The results prove that parallel CONWIP is superior over shared CONWIP in terms of the average flow time per part. If categorical dispatch rules are employed, parallel CONWIP outperforms shared CONWIP in terms of service level. With high card count, parallel CONWIP generally produces lower bottleneck utilizations while maintaining a low average flow time per part than shared CONWIP.

Keywords: push system, pull system, CONWIP, machine breakdown, multiple product types

Acknowledgement: The authors wish to acknowledge Universiti Sains Malaysia for the full financial support through the short term grant no. 6039031 and the Postgraduate Fellowship Scheme.

1 INTRODUCTION

In recent years, the use of pull systems has become increasingly prevalent in many industries to reduce WIP and other lean wastes directly or indirectly. There are many forms of pull systems, such as the single-stage kanban system, which has been thoroughly investigated through mathematical formulations and simulation studies. There are also reported industrial case studies of its applications over a long period. In comparison, constant WIP (CONWIP) systems are relatively less explored, particularly its behavioral peculiarity against emergent production challenges. One challenge faced by the CONWIP system is the production synchronization of multiple part types (Spearman, 1990). Ryan and Vorasayan (2005) proposed a method of allocating cards to each part type present in a part family, as opposed to sharing cards among part types. The simulation study revealed superior performance in terms of service level, but several flaws are overlooked. Finished goods inventory for each part type, irrespective of the consumption rate, yields higher holding cost. Setting WIP levels for each part type is unfavorable in a non-steady state environment.

Another challenge faced in CONWIP system studies is the inclusion of machine breakdown in performance comparison. With the exception of the study of Ozbayrak, Cagil, and Kubat (2004) where there are 35 part types, most literature on CONWIP systems that include breakdown only considers one part type. In reality, a particular product family employing CONWIP systems can be made up of several part types. Thus, although the superiority of the CONWIP system in the presence of breakdown is clear, this fact may not hold true if several part types are considered.

This paper compares the performance of one push system and two pull systems in a $D/D/1/\infty/\infty$ queue of a serial production line, with breakdown following a distribution and no setup. The two pull systems described are the shared and parallel CONWIP systems. The paper is organized as follows. Sections 2 provide literature a review on the push and pull systems, CONWIP system, and machine breakdown. Section 3 and 4 describes the methodology and model construction of the study respectively. Section 5 highlights the results obtained from the study. Section 6 provides the analysis of the results. Section 7 concludes the paper.

2 LITERATURE REVIEW

Production control mechanisms can be divided into two categories: push system and pull system. In the push system, production is initiated when demand is scheduled to individual workstations and parts are available for processing. In the pull system, production is initiated when finished goods/WIP inventory are withdrawn and parts are available for replenishment. The push system is more commonly employed in industries because it emerged long before the pull system did. However, although the push system has shown relative success in industries, errors in demand forecasting may cause excess/deficient finished goods/WIP inventory, and overutilization/underutilization of capacity in meeting the actual demand. Either way, unnecessary costs may be accrued. Several production planning tools associated with the push system are material requirement planning (MRP) and manufacturing resource planning (MRP II).

Around the same time that MRP and MRP II emerged, Japan was facing ordeals in developing its automotive industry. Lean manufacturing emerged as a solution, where complete waste elimination is the goal of the industry. One core principle of lean manufacturing is just-in-time. The pull system stems from this principle. A study conducted in the 1990s revealed the superiority of the pull over the push system. American and European industries were thus drawn to the pull system and its potential benefits. Womack, Jones, and Roos (1990) provided an excellent review on this shift in production control preference. Several notable benefits realized with the pull system include usage of actual demand in production and consideration of capacity utilization in setting WIP levels. In the long run, inventory costs are lowered.

One form of pull system that has gained wide acceptance from both academic and industry perspectives is the CONWIP system. CONWIP system was introduced by Spearman, Woodruff, and Hopp (1990). The basis of CONWIP system operation is that in order for parts to be admitted into the line, each part container should be attached with a card. When a container is consumed at the end of the line, the card attached to it is returned to the beginning of the line, and subsequently attached to a designated part container before being readmitted to the line. Thus, the consumed container is replenished upon completion of the designated part. Huang, Wang, and Ip (1998) compiled the notable benefits of the CONWIP system from previous studies.

One common ground in many studies on the CONWIP system is the determination of system parameter value(s) for a desired performance level. Hopp and Roof (1998) established a method known as statistical throughput control used to set WIP levels in production and assembly lines employing the CONWIP system based on a desired throughput level. Cao and Chen (2005) developed and solved a mathematical model to obtain optimal part assignment, part sequence, and lot size in production and assembly lines employing the CONWIP system. The performance measures used are setup time and work load balance. Framinan, Ruiz-Usano, and Leisten (2000) examined the effect of different dispatch rules in a five-station CONWIP system flow shop via simulation. The performance measures used are flow time, WIP level, and throughput.

Studies that compare production control mechanisms are also immense in quantity. Enns and Rogers (2008) compared the push and CONWIP systems in a simple, balanced production line. Mathematical modeling and simulation are used to evaluate the tradeoffs between throughput and inventory. Geraghty and Heavey (2004) compared hybrid push/pull and CONWIP systems via simulation. The conditions yielding optimal inventory and safety stock for the hybrid system are investigated. These conditions are in turn applied to the CONWIP system and their performances are compared. Khojasteh-Ghamari (2009) compared the kanban and CONWIP systems applied in an assembly system in terms of the WIP level and throughput. Framinan, Gonzalez, and Ruiz-Usano (2003) have highlighted other notable comparisons.

Most studies on the CONWIP system are solely from the academic perspective. This fact does not diminish the effectiveness of the CONWIP system because there are studies that discuss its industrial application, although few and far between. Spearman, Hopp, and Woodruff (1989) reported a CONWIP system in the process of installation by a large computer manufacturer. The background of the concept introduced in the factory and specific implementation issues such as throughput monitoring is discussed. Gilland (2002) introduced the CONWIP system in an Intel microprocessor factory. Several wafer release policies in cases with single and multiple bottlenecks are further analyzed via simulation.

The existence of multiple part types in production has been dealt with in several ways. In many cases, part types originating from a given product family possesses some form of common ground in production. This situation impedes independent operational control over each part type. Duri, Frein, and Di Mascolo (1995) investigated the presence of two part types in a facility operating via the kanban system. The machine utilization and mean waiting time are derived under varying processing order of products. Kenne, Boukas, and Gharbi (2003) investigated an FMS producing two part types via mathematical formulation. Numerical examples revealed the necessary level of WIP for each part type under varying breakdown and repair rates. Both these studies reveal the importance of addressing the influence of multiple part types in production; failure to do so may accrue unnecessary costs (Gharbi and Kenne, 2003).

Smalley (2009) proposed a method where part types are grouped based on their demand pattern. Each group operates via a common production control mechanism, but differs in the setting. Despite consideration of additional factors is required, the primary aim is to deal with environments where demand fluctuates. Although the proposed method describes works in an environment employing the kanban system, the method can be adopted in the CONWIP system. Most recently, Prakash, Chong, Mustafa, and Chin (2011) described the workings of such a system (termed parallel CONWIP system) in a three-product shared-machine facility with admittance of reworked parts. The results reveal the superiority of the said system. Some studies involve deciding on the WIP levels for various part types present. Ryan and Vorasayan (2005) allocated WIPs for each part type. Spearman (1990) as well as Ryan and Vorasayan (2005) provided two extremities in the allocation of WIP in a CONWIP system. In the former, the total WIP is cited; in the latter, the WIP for each part type is cited. In both cases, the WIP level set is independent of the demand pattern for each part type. A compromised solution meeting in the middle of these extremities is maintaining the WIP level for part types with frequent demand.

Many studies on the performance of shop floor assume 100% reliability in machines, which is far from reality. In fact, failure is a characteristic possessed by all entities on the shop floor related to production. Machine failure behavior is typically characterized by a bathtub curve (Lewis and Chen, 1994). With this behavior, machine failure, or rather, machine breakdown, may be represented by a distribution. Two parameters commonly addressed are the mean-time-to-failure (MTTF) and mean-time-to-repair (MTTR).

The impact of machine breakdown has been addressed in both push (Chung, 2003; Wazed, Ahmed, Yusoff, 2010) and CONWIP (Graves and Milne, 1997; Ozbayrak Cagil, and Kubat, 2004) systems. Several alternatives for dealing with machine breakdown have been proposed. Abboud (2001) compiled these alternatives from previous studies. Production in excess of demand is the essence of many of these alternatives. The difference between them lies in the point at which production stops and resumes. However, the main goal of any method employed is to prevent the machine prone to breakdown from becoming the bottleneck. The most well-known alternative in dealing with machine breakdown is preventive maintenance.

3 METHODOLOGY

This study is approached using discrete-event simulation. Control parameters selected are commonly used by management to influence the performances of production. The experimental design approach used is a full-factorial experimental design, whereby experiments are performed on all possible combinations of discrete values prescribed in the control parameters. The performance measures are collected after each simulation run. The model is constructed in WITNESS® 2008.

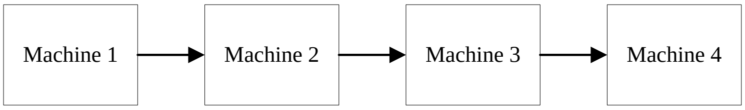
A multi-factor ANOVA is performed to determine the effects of the control parameters and their interactions on each performance measure. In addition, response surface methodology (RSM) based on Box and Wilson (1951), relates any significant effect found with the specified performance measure. A second-order polynomial that yields a high value of R^2 is selected in depicting the behavior of a model. Certain strategies are employed to make comparison between models possible. In most cases, two independent variables are present in an equation. Recommended by Montgomery (1997), the analysis can be simplified by setting one variable as constant.

Banks, Carson II, Nelson, and Nicol (2005) reported that one method of model verification is to examine model output for reasonableness. The behavior of each model is tested by changing the control values of one parameter at a time. The models are verified in two methods. First, the models are made deterministic and the results obtained from the initial simulation run are compared with ones generated from manual calculation. The models are fine-tuned until both results are matched. Second, the performances of the models are observed and compared with established general system behavior in found Groover (1987).

Model Description

The simulation model consists of four automated machines in a series, processing two categories of parts, namely high runner (HR) and low runner (LR). The model emulates a D/D/1/∞/∞ input queue with saturated demand. The inter-arrival and service times are deterministic for a single production line, with unlimited queue capacity and infinite population of potential arrivals. The HR consists of parts with frequent demand, whereas the LR comprises parts with infrequent demand. Each category has two part types: in total, HR_1, HR_2, LR_1, and LR_2. Each machine is separated by a buffer. Parts arrive at an inter-arrival time of eight hours. To avoid excessive WIP build-up, an additional constraint is introduced such that the total number of parts to be processed upon each inter-arrival time is fixed, as opposed to a fixed inter-arrival quantity. Each batch is made up of LS parts containing either HR or LR. The distribution of HR and LR at each arrival is defined by HRLR, which is the ratio of HR to LR. All parts proceed via the same route asynchronously, as shown in Figure 1.

Figure 1: Processing route of parts



Each machine follows a deterministic processing random outage case (DPRO), where processing times are deterministic, but machines are subjected to breakdown and repair following exponential distribution. The MTTF is commonly adopted as a parameter of choice in depicting breakdown; however, it has the flaw of breakdown occurring even when the machine is idle. This effect is magnified in the CONWIP system, where Spearman, Woodruff, Hopp (1990) highlighted that machines are periodically idle. To enable a fair mode of comparison against the push system, the parameter representing breakdown is the number of operations, BC. This variable indicates that a machine experiences breakdown as soon as BC operations are completed. All machines are assumed to be equally unreliable. Thus, MTTR has a fixed value.

Machine processing times do not follow a distribution primarily because such systems may have shifting bottlenecks, that is, different bottlenecks at each replication. Therefore, a deterministic processing time is more suitable because bottleneck utilization is of interest. Three production control strategies discussed in this paper include the push (P), shared CONWIP (SC), and parallel CONWIP (PC) systems. In the P system, batches are pushed to an available machine as long as a preceding buffer is not empty. In the SC system, one batch is admitted into the line when one CONWIP card is available. KC cards are shared between HR and LR batches. In the PC system, one HR batch is admitted into the line when one HR CONWIP card is available. The same applies to LR part batches. KCHR cards are shared between HR batches and KCLR between LR batches. Prakash, Chong, Mustafa, and Chin (2011) depicted diagrams describing the operation of each system.

Dispatch rules dynamically rank queues by computing batch priority indices (Bhaskaran and Pinedo, 1992). In this study, three dispatch rules are considered: first-in-first-out (FIFO), HR-LR (HL), and LR-HR (LH). FIFO is based on the arrival time, where batches of earlier arrival times are given higher priority. HL prioritizes HR over LR, and if more than one batch of a category is present, the FIFO rule is in place. LH follows HL, except that LR is prioritized over HR. The HL and LH rules apply only to the PC system due to the distinction made within the system. It is run for one shift a day, five days a week for over 10 weeks (1 440 000 s) with a warm-up period of two weeks (288 000 s). Table 1 summarizes the variable values of each control parameter used in the simulation.

Table 1: Control parameter variable values

Control parameter	Variable value
LS	50, 100, 150
HRLR	0.3, 0.5, 0.7
BC	10, 25, 40
KC	6, 10, 14
KCHR	3, 5, 7
KCLR	3, 5, 7

The performance measures of interest are the HR service level, LR service level, throughput, average flow time per part, and bottleneck utilization. These performance measures are selected because of their prevalent use in the industry and scientific studies. The WIP is not one of the performance measures of interest because according to Little (1961), WIP level can be estimated once throughput and average flow time per part are known.

When all machines are 100% reliable in a push system, the lead time of HR providing a 95% service level is set as the HR lead time for all models. The same applies for LR. This method of setting due dates is known as the endogenous due date setting (Cheng and Gupta, 1989), and takes into account information on an arriving job. In P-FIFO, a part is committed to the line as soon as it is pushed to the immediate upstream buffer of the first machine. In the SC-FIFO and PC variants, a part is committed to the line as soon as a card is attached to the order. The following annotations and mathematical expressions define each performance measure:

LT_{HR} = lead time of HR
LT_{LR} = lead time of LR
a_{iHR} = time when part i of category HR arrives
b_{iHR} = time when part i of category HR is committed
c_{iHR} = time when part i of category HR is shipped
d_{iHR} = due date of part i of category HR
a_{iLR} = time when part i of category LR arrives
b_{iLR} = time when part i of category LR is committed
c_{iLR} = time when part i of category LR is shipped
d_{iLR} = due date of part i of category LR
i = 1, 2, 3….
T_{wu} = warm-up period

Let f_{HR}=0, f_{LR}=0, g_{HR}=0 and g_{LR}=0 at the start of each simulation run

if $t \geq T_{wu}$
 $d_{iHR} = a_{iHR} + LT_{HR}$
 $d_{iLR} = a_{iLR} + LT_{LR}$
 if $c_{iHR} \leq d_{iHR}$
  $f_{HR} = f_{HR+1}$
 else
  $g_{HR} = g_{HR+1}$
 endif
 if $c_{iLR} \leq d_{iLR}$,
  $f_{LR} = f_{LR+1}$
 else
  $g_{LR} = g_{LR+1}$
 endif
endif

HR service level = $\frac{100 f_{HR}}{f_{HR} + g_{HR}}$

LR service level = $\frac{100 f_{LR}}{f_{LR} + g_{LR}}$

$$\text{Throughput} = \frac{f_{HR} + g_{HR} + f_{LR} + g_{LR}}{1440000 - 288000} = \frac{f_{HR} + g_{HR} + f_{LR} + g_{LR}}{1152000}$$

$$\text{Average flow time per part} = \frac{\sum_{i=1}^{f_{HR}+g_{HR}} (c_{i\,HR} - b_{i\,HR}) + \sum_{i=1}^{f_{LR}+g_{LR}} (c_{i\,LR} - b_{i\,LR})}{f_{HR} + g_{HR} + f_{LR} + g_{LR}}$$

$$\text{Bottleneck utilization} \approx \frac{100 \times 100(f_{HR} + g_{HR} + f_{LR} + g_{LR})}{1440000 - 288000} = \frac{f_{HR} + g_{HR} + f_{LR} + g_{LR}}{115.2}$$

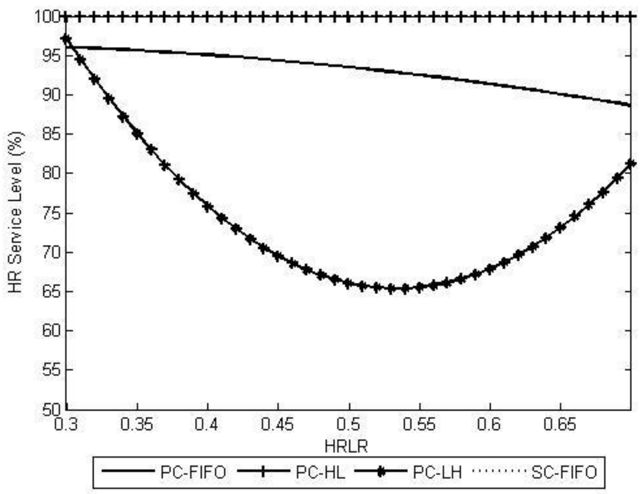
4 RESULTS

The ANOVA results of the three production controls are explained in this section. In all PC variants, all main factors pose effects on the performance measures of this study. Interactions of any two main effects are also significant. However, in PC-HL, interaction of KCLR with any remaining factors poses no effect on the performance measures of this study and LR approximates a push system behavior. A similar observation is obtained with PC-LH where HR approximates a push system behavior. In addition, interactions of higher order reveal negligible effect on the performance measures.

With regards to SC-FIFO, all performance measure in the study reveals no significant changes with respect to HRLR. This is due to the absence of assignment of cards between HR and LR orders. Likewise, within the context of interaction, interactions of HRLR with any remaining factors poses no significant effect on the performance measures of this study.

In P-FIFO the performance measures of this study are independent of changes in HRLR. This is consistent with a push system behavior where the only determinant is the lot size, and manipulation of this variable brings about significant changes in the aforementioned performance measures. However, the analysis also reveals that changes in BC bring no significant effect to HR and LR service level as well as average flow time per part. This is due to the sufficiently high WIP levels within the system such that the effect of breakdown is suppressed. As for interaction between factors, only changes in both BC and LS impose significant effect in the performance measures of this study.

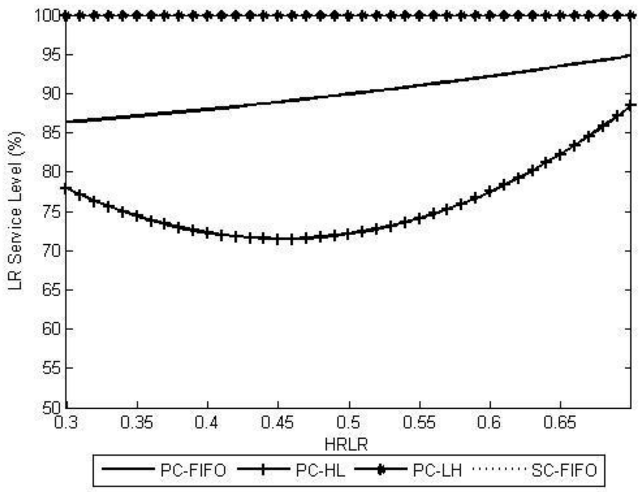
Figure 2: HR service level vs. HRLR for SC-FIFO, PC-FIFO, PC-HL, and PC-LH.



SC-FIFO overlaps with PC-HL

The HR service level of PC-HL is as good as that of SC-FIFO (Figure 2), maintaining 100% HR service level irrespective of HRLR. For PC-FIFO, although the HR service level is adequately high, it is still lower than that of SC-FIFO, and exhibits a decrease with HRLR. PC-LH has the lowest HR service level among all systems, and exhibits a minimum at approximately HRLR=0.5.

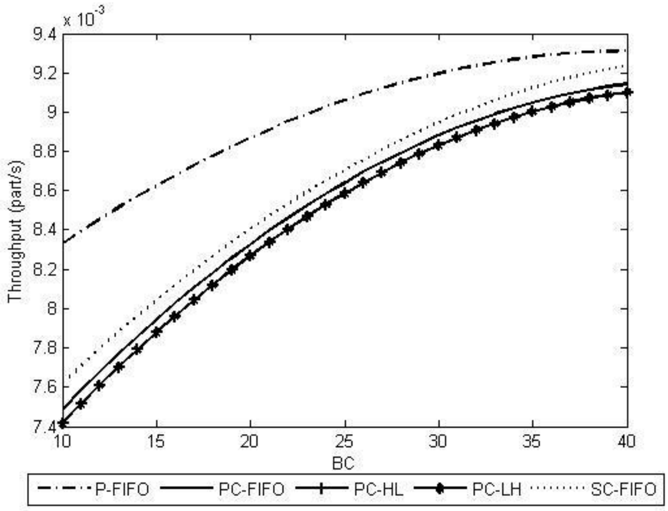
Figure 3: LR service level vs. HRLR for SC-FIFO, PC-FIFO, PC-HL, and PC-LH.



SC-FIFO overlaps with PC-LH

The LR service level of PC-LH is as good as that of SC-FIFO (Figure 3), maintaining 100% HR service level irrespective of HRLR. For PC-FIFO, although the HR service level is adequately high, it is still lower than that of SC-FIFO, and exhibits an increase with HRLR. PC-HL has the lowest HR service level among all systems, and exhibits a minimum at approximately HRLR=0.45. LR service level reflects a similar behavior to that of HR service level.

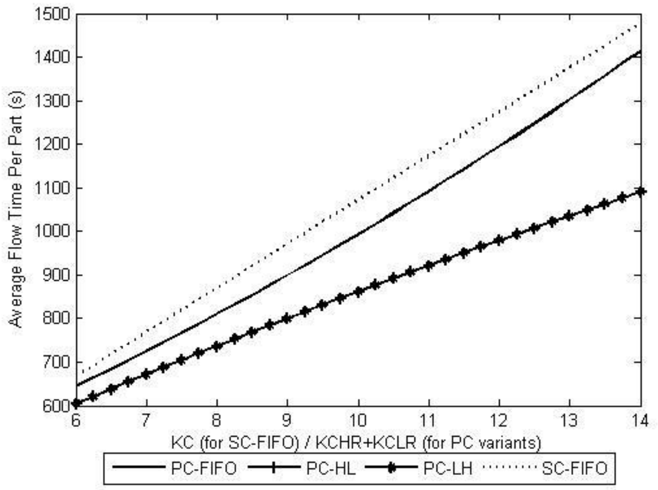
Figure 4: Throughput vs. BC for P-FIFO, PC-FIFO, PC-HL, PC-LH, and SC-FIFO



PC-HL overlaps with PC-LH

The throughput of P-FIFO is the highest compared to remaining systems (Figure 4). The throughput of PC-FIFO is slightly lower than SC-FIFO. With the HL and LH dispatch rules, throughput is suppressed even further. Throughput of all systems shows a decrease with increasing BC.

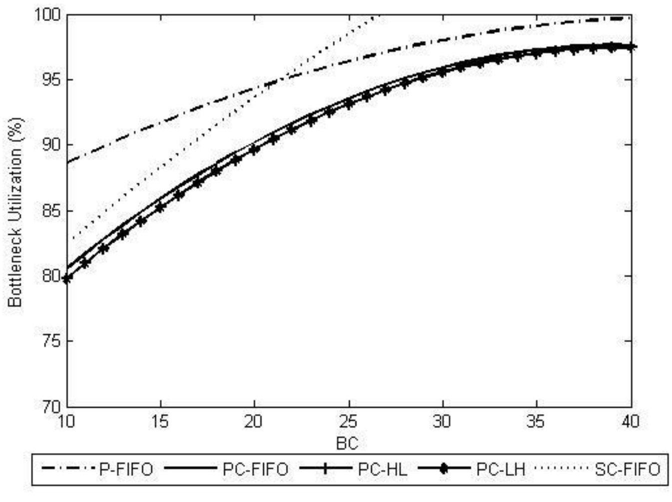
Figure 5: Average flow time per part vs. KC / KCHR+KCLR for SC-FIFO, PC-FIFO, PC-HL, and PC-LH.



PC-HL overlaps with PC-LH

The average flow time per part of SC-FIFO is the highest compared to remaining PC variants (Figure 5). PC-FIFO is slightly lower than this and the presence of HL and LH loading rules suppress the average flow time per part even further. The average flow time per part of all systems show an increase with increase in total number of CONWIP cards.

Figure 6: Bottleneck utilization vs. BC for SC-FIFO, PC-FIFO, PC-HL, and PC-LH.



PC-HL overlaps with PC-LH

The bottleneck utilization of all PC variants is constantly lower than those of P-FIFO and SC-FIFO (Figure 6). The bottleneck utilization of P-FIFO is the highest when BC lies between 10 and 20; beyond this range, the bottleneck utilization of SC-FIFO is the highest. The bottleneck utilization of all systems exhibits an increase with increasing BC.

5 DISCUSSION

In essence, between the SC and P systems, a larger proportion of parts in the former are able to meet the due date set while maintaining a low WIP level. This finding is consistent with that of JodlBauer and Huber (2008). The primary reason for this behavior is the limit placed on WIP in SC-FIFO. In P-FIFO, WIP accumulates at the immediate upstream buffer of the bottleneck; with higher breakdown rate, WIP accumulation is higher. This finding highlights one benefit of the SC over the P systems; in the presence of breakdown, the admittance of fresh parts is suppressed due to the absence of free cards. Although machine breakdown is in fact an unpredictable event, the occurrence of breakdown naturally controls fresh parts from being admitted into the line, in accordance with the number of cards present. The absence of this control mechanism in the event of breakdown, as in the P system, affects the service level.

The HR service levels in decreasing order are SC-FIFO and PC-HL, PC-FIFO and P-FIFO, and PC-LH. The allocation of cards between HR and LR may cause instances when cards of a particular category are not available for attachment to corresponding orders (noted among PC variants in Figure 2). This void can be manipulated to favor a given category if priority is given to parts of that category. However, this void does not occur with SC-FIFO. On a different note, with PC-LH, the HR service level is the lowest and exhibits a minimum. With increased HRLR up to the minimum point, the quantity of HR increases. The LH dispatch rule is sufficient to suppress the effect of increased HR volume. However, beyond this point, the HR service level increases, and as at any given instance, LR can be absent from a queue. Although PC-FIFO does not perform as well as SC-FIFO in terms of the service level, the practical simplicity of the categorical dispatch rules are sufficient to elevate the service level.

The LR service levels in decreasing order are SC-FIFO and PC-LH, PC-FIFO and P-FIFO, and PC-HL. Between the SC and PC systems, the former maintains its behavior as before, whereas the behavior of PC variants is the inverse. In Figure 3, the minimum point in the LR service level of PC-HL is explained as follows. At HRLR values lower than that at the minimum point, the effect of the HL dispatch rule coupled with increased HR volume has a negative effect on the LR service level. Beyond this point, although LR orders have the two said factors working against it, the LR service level increases. With a smaller volume of LR, there is lesser chance for a given LR order not to meet the specified due date.

The throughputs in decreasing order are P-FIFO, SC-FIFO, PC-FIFO, and PC-HL and PC-LH. Figure 4 shows that the throughput of P-FIFO remains highest as parts are constantly admitted into the line irrespective of downstream needs. P-FIFO, despite producing more parts, still yields a lower service level. Between the SC and PC systems, the WIP levels are limited by the number of cards present, hence the lower throughput. Throughputs of PC variants are lower than those of SC-FIFO, albeit only a small difference. This slight difference in throughput is sufficient evidence of the possible difference in the net WIP present in the line. This difference also indicates that although PC variants have the same total number of cards as in SC-FIFO, PC variants have more instances when cards are not in use, hence the lower service level than that in SC-FIFO.

The SC system has a lower and constant average flow time per part compared with the P system. This finding is consistent with that of Spearman and Zazanis (1992), who also compared the behavior of the SC and P systems. The average flow times per part in decreasing order are P-FIFO, SC-FIFO, PC-FIFO, and PC-HL and PC-LH. Any given part in P-FIFO spends a large proportion of time waiting for processing to begin. A lower average flow time per part corresponds to a larger lot size, and a larger lot size comes with increased WIP level. On the other hand, SC-FIFO exhibits a constant average flow time per part. With a larger lot size, each batch spends a longer time in processing, thereby delaying the cards from being freed and limiting the net WIP in the line. Lesser WIP queues indicate lesser waiting time.

Between the SC and PC systems, the summation KCHR and KCLR in PC variants at any given instance corresponds to the equivalent KC in SC-FIFO. In Figure 5, due to the absence of allocation between HR and LR in SC-FIFO, both categories have equal chances of obtaining a card. However, in PC-FIFO, a larger HRLR increases the average flow time per part, as more orders need to wait for HR cards in the line to be freed. This phenomenon causes the range of average flow time per part in PC-FIFO to be larger than that in SC-FIFO. This larger range also accounts for the lower average flow time per part in PC-FIFO. With HL and LH dispatch rules, this range is increased even further, constituting a lower average flow time per part. Although the flow time of the PC variant diminishes its predictability, as in the report of Spearman (1990) where an SC system flow time remains effectively constant, this feature can be seen as an advantage.

For PC variants, in the event of an influx of parts of a given category, categorical dispatch rules or introduction of additional cards favoring that category can be used to cater to this sudden change. Another option is the conversion of cards from an opposing category to the category of the influxed parts. In SC-FIFO, the only way to meet this requirement is by the introduction of additional cards, which increases the net WIP, hence the average flow time per part. From a different perspective, although PC variants generally yield a lower service level than SC-FIFO, its shorter average flow time per part can be taken advantage of. With advanced knowledge of flow time, a card can be freed earlier in a PC environment, thereby allowing jobs an early start.

With knowledge on the average flow time per part and throughput, the WIP level in each system can be deduced in accordance with Little's law (Little, 1961). The WIP levels in decreasing order are P-FIFO, SC-FIFO, PC-FIFO, and PC-HL and PC-LH. As pointed out by Karmarkar (1986), the number of cards used in a pull system does not indicate the net WIP present, but rather, sets an upper limit on the WIP level. This theory applies to the CONWIP system as well. As pointed out earlier, between the SC and PC systems, there are more instances in PC variants when cards are not in use. In the event of breakdown in a PC environment, these unused cards can be kept as reserves to allow for machine repair. In SC-FIFO, the reservation of WIP is more difficult to implement if all cards are present in the line.

The bottleneck utilizations in increasing order are PC-HL and PC-LH, PC-FIFO, and SC-FIFO and P-FIFO coinciding at a BC of approximately 25 (Figure 6). At smaller BC values, P-FIFO exhibits the highest bottleneck utilization. At higher BC values, SC-FIFO exhibits the highest bottleneck utilization. In SC-FIFO, with higher breakdown rate, fresh parts are controlled from entering the line, hence the lower bottleneck utilization. As aforementioned, PC variants exhibit the lowest bottleneck utilization due to the effect of allocation between HR and LR. The lower bottleneck utilization is sufficient indication that the introduction of additional cards to the system is still a valid option. Although higher utilization is generally more desirable for investment justification purposes, it may appear as a disadvantage in the event of influx of parts. In such a case, as in SC-FIFO, the introduction of additional cards may yield a lower service level with increased net WIP (where average flow time per part may increase), but this may not be possible as the bottleneck is at capacity.

Thus, an additional experiment is carried out to compare the performance of PC-FIFO and SC-FIFO at common values of bottleneck utilization. These common values are obtained are as follows: the number of cards in PC-FIFO is increased such that its behavior with respect to BC approximates that of SC-FIFO. The HR (Figure 7) and LR (Figure 8) service levels of the new PC-FIFO follows that of its predecessor. However, the average flow time per part of PC-FIFO increases at a lower gradient than SC-FIFO, as shown in Figure 9. Therefore, at a given value of bottleneck utilization and in systems with high number of cards, PC-FIFO is advantageous over SC-FIFO in terms of average flow time per part.

Figure 7: HR service level vs. HRLR for PC-FIFO (with increased bottleneck utilization) and SC-FIFO.

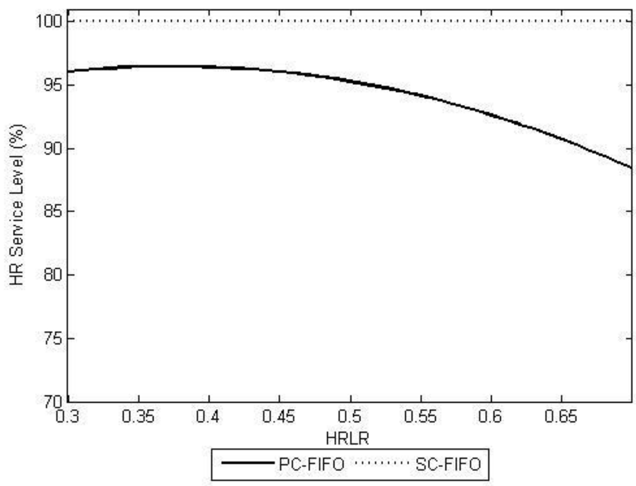


Figure 8: LR service level vs. HRLR for PC-FIFO (with increased bottleneck utilization) and SC-FIFO.

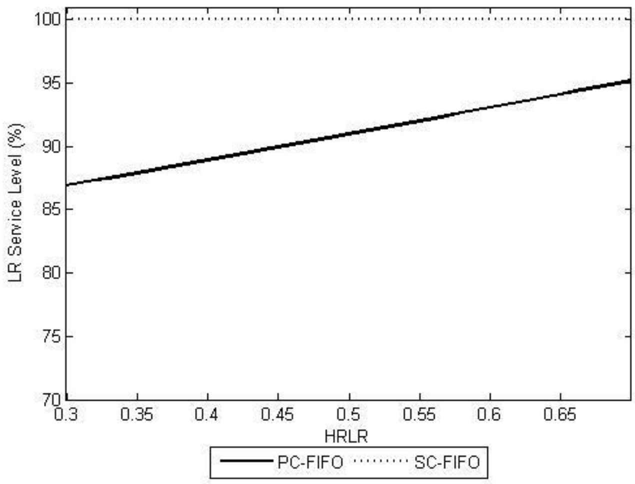
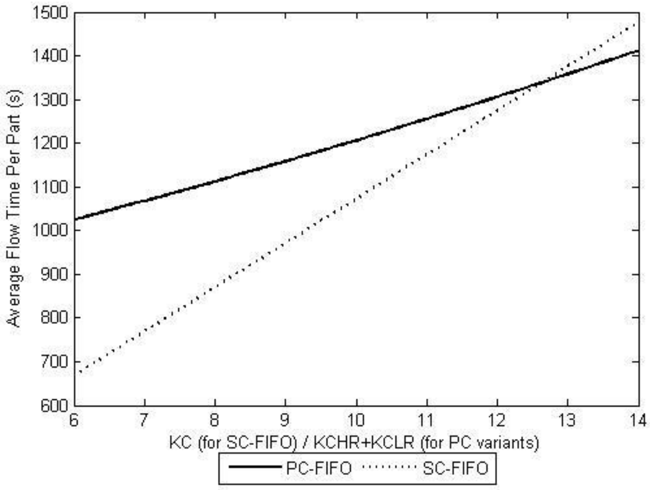


Figure 9: Average flow time per part vs. KC / KCHR+KCLR for PC-FIFO (with increased bottleneck utilization) and SC-FIFO.



6 CONCLUSION AND FURTHER RESEARCH

This paper compares the performance of the P, SC, and PC systems in the presence of breakdown. SC has the advantage of higher HR and LR service levels over PC systems. However, PC systems are superior over the SC system in terms of lower average flow time per part and lower bottleneck utilization. The P system is superior over the SC and PC systems only in terms of throughput.

The study is highly beneficial for flow shop production facilities intended to adopt high mix by providing alternative production planning and control systems notably using CONWIP. In such production environment, demands are obtained from different customers, consequently product life cycles are short and production volumes are significantly varied in time. The process generally consists of operations with heterogeneous groups of machines. Machine breakdown is common due to high speed processing as well as constant product change. In particular, the results reveal the advantages of parallel CONWIP systems in terms of average flow time per part irrespective of the fluctuating demand ratio of high runners and low runners. In comparison, a shared CONWIP system will require part sequencing in the backlog list to effectively accommodate sudden changes in these demands. In addition, a more volatile demand change may require rigorous shuffling of part sequencing in the backlog list. Aside from this, a recalculation of CONWIP card may be required. In parallel CONWIP systems, part sequencing is not required as categories of parts are predefined. Thus the backlog list may focus on efforts in determining suitable number of cards dedicated for high runner and low runner. In the long run, the burden placed on the scheduling system is alleviated.

The PC system is relatively new and thus possesses many unexplored options. The effect of categorical dispatch rules in the PC system shows better performance in terms of category service level and average flow time per part. However, the possibility of other dispatch rules commonly adopted in production floors may change the behavior of these performance measures. Studies on the effect of the machine setup on such a system can also be interesting because the setup time can have a wide range of values, depending on the industry of application. A larger setup time may constitute a larger deviation from the aforementioned behavior. The possibility of increasing the service level by releasing cards earlier than the completion of processing is also of interest. One final notable aspect of interest is the implementation procedure of such systems and how it differs from conventional CONWIP system implementation.

REFERENCES

Abboud, N. E. (2011). A discrete-time Markov production-inventory model with machine breakdowns. *Computers and Industrial Engineering*, *39*, 95-107.

Banks, J., Carson II, J. S., Nelson, B. L., & Nicol, D. M. (2005). *Discrete-Event System Simulation*. Englewood Cliffs, NJ: Prentice-Hall.

Bhaskaran, K., & Pinedo, M. (1992). *Handbook of Industrial Engineering*, edited by G. Salvendy. New York, NY: John Wiley.

Box, G. E. P., & Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society, Series B (Methodological)*, *13*(1), 1-45.

Cao, D., & Chen, M. (2005). A mixed integer programming model for a two line CONWIP based production and assembly system. *International Journal of Production Economics*, *95*, 317-326.

Cheng, T. C. E., & Gupta, M. C. (1989). Survey of scheduling research involving due date determination decisions. *European Journal of Operational Research*, *38*, 156-166.

Chung, K. J. (2003). Approximations to production lot sizing with machine breakdowns. *Computers and Operations Research*, *30*, 1499-1507.

Duri, C., Frein, Y., & Di Mascolo, M. (1995). Performance evaluation of kanban multiple-product production systems. *Emerging Technologies and Factory Automation*, *3*, 557-566.

Enns, S. T., & Rogers, P. (2008). *Clarifying CONWIP versus push system behavior using simulation.* Proceedings of winter simulation conference, Miami, Florida, USA. Piscataway, New Jersey: IEEE Press, 1867-1872.

Framinan, J. M., Ruiz-Usano, R., & Leisten, R. (2000). Input control and dispatching rules in a dynamic CONWIP flow-shop. *International Journal of Production Research*, *38*(18), 4589-4598.

Framinan, J. M., Gonzalez, P. L., & Ruiz-Usano, R. (2003). The CONWIP production control system: review and research issues. *Production Planning and Control*, *14*, 255-265.

Geraghty, J., & Heavey, C. (2004). A comparison of hybrid push/pull and CONWIP/pull production inventory control policies. *International Journal of Production Economics*, *91*(1), 75-90.

Gharbi, A., & Kenne, J. P. (2003). Optimal production control problem in stochastic multiple-product multiple-machine manufacturing systems. *IIE Transactions*, *35*, 941-952.

Gilland, W. (2002). A simulation study comparing performance of CONWIP and bottleneck-based release rules. *Production Planning and Control*, *13*(2), 211–219.

Graves, R. J., & Milne, J. R. (1997). A new method for order release. *Production Planning and Control*, *8*(4), 332-342.

Groover, M. P. (1987). *Automation, Production Systems and Computer Integrated Manufacturing*. Englewood Cliffs, NJ: Prentice-Hall.

Hopp, W. J., & Roof, M. L. (1998). Setting WIP levels with Statistical Throughput Control (STC) in CONWIP production lines. *International Journal of Production Research*, *36*, 867-882.

Huang, M., Wang, D., & Ip, W. H. (1998). Simulation study of CONWIP for a cold rolling plant. *International Journal of Production Economics*, *54*(2), 257-266.

Jodlbauer, H., & Huber, A. (2007). Service-level performance of MRP, Kanban, CONWIP and DBR due to parameter stability and environmental robustness. *International Journal of Production Research*, *46*(8) 2179-2195.

Karmarkar, U. S. (1986). *Kanban systems*. Working Paper Series No.QM8612, Center for Manufacturing and Operations Management, The Graduate School of Management, The University of Rochester.

Kenne, J. P., Boukas, E. K., & Gharbi, A. (2003). Control of production and corrective maintenance rates in a multiple-machine, multiple-product manufacturing system. *Mathematical and Computer Modelling Journal*, *38*, 351–365.

Khojasteh-Ghamari, Y. (2009). A performance comparison between kanban and CONWIP controlled assembly systems. *Journal of Intelligent Manufacturing*, *20*, 751-760.

Lewis, E. E., & Chen, H. C. (1994). Load-capacity interference theory and the bathtub curve. *IEEE Trans. Reliability*, *43*, 470-476.

Little, J. (1961). A Proof Of The Queueing Formula L = λW. *Operations Research*, *9*(2), 383-387.

Montgomery, D. C. (1997). Response surface methods and other approaches to process optimization. In: D. C. Montgomery. *Design and analysis of experiments* (pp. 427–510). New York: John Wiley and Sons.

Ozbayrak, M., Cagil, G., & Kubat, C. (2004). How successfully does JIT handle machine breakdowns in an automated manufacturing system? *Journal of Manufacturing Technology Management*, *15*(6), 479-494.

Prakash, J., Chong, M. Y., Mustafa, S. A., & Chin, J. F. (2011). Parallel CONWIP for a high-mix multi-stage production system with entrance of rework. *Journal of Production Research and Management*, *1*(2), 1-23.

Ryan, S. M., & Vorasayan, J. (2005). Allocating work in process in a multiple-product CONWIP system with lost sales. *International Journal of Production Research*, *43*, 223-246.

Smalley, A. (2009). *Creating Level Pull*. Cambridge, MA: Lean Enterprise Institute.

Spearman, M. L., Hopp, W. J., & Woodruff, D. L. (1989). A hierarchical control architecture for Constant Work-in-Process (CONWIP) production systems. *Journal of Manufacturing and Operations Management*, *2*, 147-171.

Spearman, M. L., Woodruff, D. L., & Hopp, W. J. (1990). CONWIP: a pull alternative to kanban. *International Journal of Production Research*, *28*(5), 879-894.

Spearman, M. L., & Zazanis, M. A. (1992). Push and pull production systems: Issues and comparisons. *Operational Research*, *40*, 521-532.

Wazed, M. A., Ahmed, S., & Yusoff, N. (2010). Impacts of common processes in multistage production system under machine breakdown and quality uncertainties. *African Journal of Business Management*, *4*(6), 979-986.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The Machine That Changed the World: The Story of Lean Production*. New York: Harper Collins Publishers.